Complementary LC- and GC- Mass Spectrometry Techniques Provide Broader Coverage of the Metabolome

Cross-Platform Metabolomics Data Analysis Combining AB SCIEX LC/MS, LECO GC/MS, and Genedata Software

Baljit Ubhi[1], Jeff Patrick[2], David Alonso[2], and Joe Shambaugh[3]
[1]AB SCIEX, Redwood City, CA, [2]LECO Corporation, St. Joseph, MI, [3]Genedata Inc., Lexington, MA.

The interaction between primary and secondary metabolism in organisms remains key to furthering our understanding of biology. Despite their obvious interconnectivity, in most metabolomics studies either primary or secondary metabolites are analyzed to reveal the metabolic response of the system to a specific perturbation. Metabolomics researchers thus often need to use both GC/MS and LC/MS technologies to provide more comprehensive coverage of the analytes in biological systems. And in order to distinguish between nearly isomeric and isobaric analytes, which is often the case for untargeted metabolomics, high-resolution mass spectrometers are the instrument of choice. High resolution also allows accurate mass measurement of pseudo-molecular ion signals and the calculation of empirical formulas, thus providing additional confidence to confirm metabolite identifications.

Here, using a well-established rat model for diabetes, obesity, and cardiovascular disease effects, we highlight the additional value provided by using both high resolution GC/MS and LC/MS analyses for untargeted metabolomics in an integrated workflow. In addition, the ability to process both data types within a single powerful software solution provides an even more comprehensive understanding of the underlying biology.

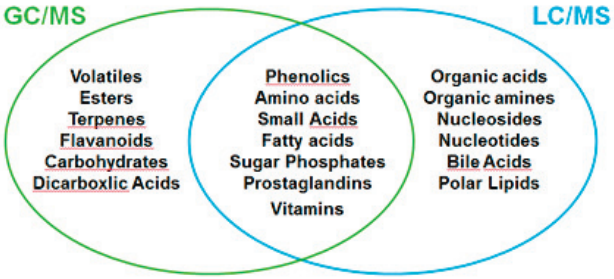
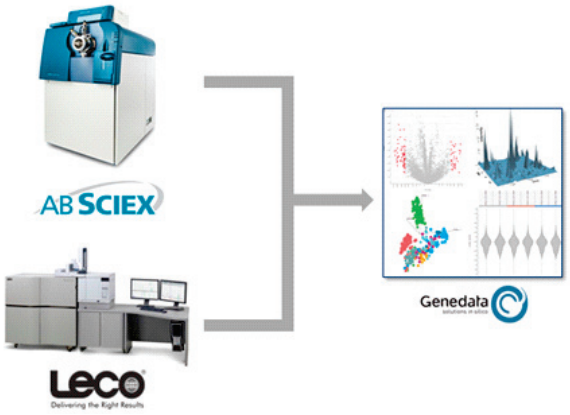


Figure 1. Combining GC/MS and LC/MS for Broad Metabolome Coverage. The two analytical techniques allow detection and quantitation of a complementary set of metabolites with some overlap, providing broader coverage of the metabolome.



Key Features of Combining LC/MS and GC/MS Technologies and Data Analysis

- High resolution accurate mass spectral data from LC and GC platforms:
  - **Pegasus® GC-HRT System** provides analysis of a wide variety of compounds, including short chain alcohols, esters, amino acids, and nucleotides
  - **TripleTOF™ 5600+ System** provides analysis of amines, semi-volatiles, lipids, and high polarity metabolites
- Combined data analysis using **Genedata Expressionist® Software** providing a comprehensive understanding of the biological system.
- Link to pathway mapping and biological interpretation.
- Extra level of confidence to your biological interpretation and validation of your results.
- Complementary LC- and GC-mass spectrometry provides broader coverage of the metabolome.

Life Science and Chemical Analysis Solutions

Materials and Methods

Sample Preparation: Plasma aliquots from Zucker rats were obtained from Bioreclamation (Westbury, NY, USA). Plasma from three strains of rats were obtained – lean, fatty, and obese phenotypes. The study used 11 lean samples, 10 obese samples, and 9 fatty samples in total. The animals were 7-9 weeks old. Animals were fed *ad libitum* chow which was 18% protein, 6% fat. The plasma was obtained from a terminal bleed and was preserved only with EDTA and maintained at -80°C. Protein was removed from the samples by the addition of 8 volumes of cold methanol. After mixing, the samples were centrifuged at 10,000 rpm in a bench-top centrifuge and the supernatant retained. The resulting solution was then portioned for analysis and aliquots dried and reconstituted as appropriate.

***TripleTOF* 5600+ LC/MS System:** Using an Agilent 1290 HPLC system and a high strength silica column (Acquity HSS T3 1.8 µm, 2.1 x 100 mm, Waters, Milford, US) @ 60°C polar metabolites from a 5 µL injection of serum were separated at a flow rate of 600 µL/min. Full scan TOFMS and MS/MS data was acquired on a *TripleTOF* 5600+ system in data dependent mode (AB SCIEX). A pooled sample, used as a QC, was also acquired every 5 samples and used to monitor data reproducibility.

***Pegasus* GC-HRT GC/MS System:** Using an Agilent 7890 GC system and an Rxi-5Sil MS column (30 m x 0.25 mm x 0.25 mm) and 5 m guard column (Restek, Bellefonte, PA, USA), metabolites from a 1 µL injection of serum were separated at a flow rate 1 ml/min (constant flow) with helium as the carrier gas. Acquisition was performed using EI (70 eV) in high resolution mode (R=25,000). The mass range acquired was 60 to 520 at 6 spectra/second. Mass calibration was performed using PFTBA. A QC sample was injected after approximately every 6th sample.

Data Processing: Mass spectrometry data from the LC/MS/MS experiments were processed using **Genedata Expressionist®** for Mass Spectrometry software. This proceeded in a stepwise fashion to first process the raw data files resulting in three data matrices, and second, to combine and interpret the data. The first step was performed using workflows which load the raw data files, remove background noise, identify peaks, cluster isotopes, and assign metabolite identities based on library search results. The result of the initial data processing workflow were data matrices consisting of the intensity values corresponding to the metabolite clusters across all chromatograms. Separate data matrices were generated for the LC/MS/MS positive and LC/MS/MS negative mode data. The GC/MS data from LECO was processed through a similar workflow and resulted in a third data matrix. In the second step, data from the three data matrices were normalized and combined together to facilitate integrated statistical analysis.

An initial QC assessment for each data matrix was performed in order to identify and remove outliers. After outlier removal, the data were normalized through Z-transformation then mapped to the original animal IDs in order to create one large data matrix to be used for statistical analysis to identify differentially expressed metabolites. Differentially expressed metabolites identified with ANOVA analysis were visualized with Principal Component Analysis and correlation network maps.

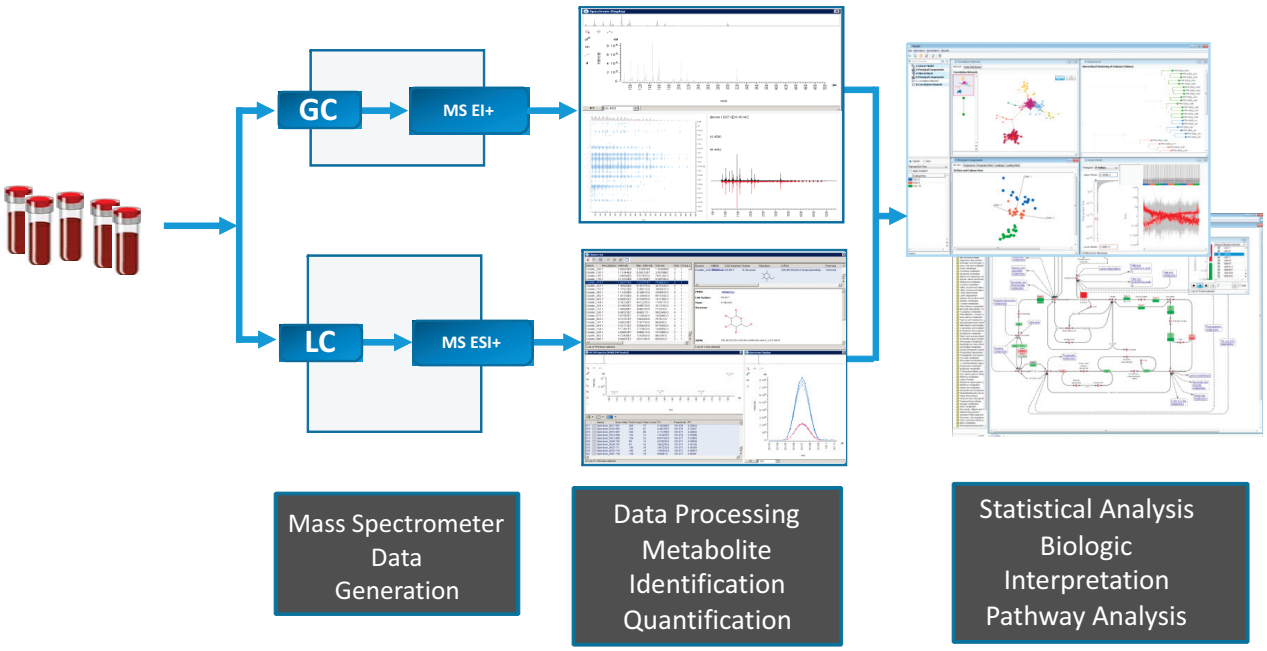


Figure 2. Workflow for Combining High Resolution Accurate Mass LC/MS and GC/MS Data with Genedata software. Bringing the LC/MS data from the AB SCIEX TripleTOF 5600+ system together with the GC/MS data from the LECO Pegasus GC/MS system with the powerful Genedata Expressionst software provides a broader coverage of analytes for global metabolomics studies.

Power of the LC/MS Data

Interrogating the LC/MS data, it was observed that there were many lipids changes amongst the three groups of rats (lean, fatty, and obese). Specifically, changes in the glycerolipids, glycerophospholipids, sphingolipids, cholesterol esters, and carnitines were detected. Significant changes amongst the bile acids were also detected between all three sample groups. Low signal intensities for the short chain acyl carnitine 4:0 was observed in the lean rats compared to the other two groups indicating an increase in the catabolism of branch chain amino acids (BCAAs) in diabetes or obesity.

Using LC/MS data alone we were only able to generate statistical models using discriminant analyses.

Power of the GC/MS Data

The majority of the metabolites found in the GC/MS data were lower molecular weight, primary metabolites, and those from core metabolic processes including amino acid metabolism, fatty acid metabolism, TCA cycle, glycolysis, and related. Specifically, changes in valine/leucine/isoleucine, all branched chain amino acids involved in fatty acid biosynthesis, alpha-hydroxycarboxylic and fatty acids, glycerol-3-phosphate, and oxo-proline were observed to be up-regulated in the fatty and obese phenotypes. Other metabolites including myoinositol and serotonin were observed to be down-regulated in the fatty and obese phenotypes. Many of these modulations correlated with pathways implied by the LC/MS analyses.

Power of Combined Technologies with Genedata

Combining the LC/MS and the GC/MS data into Genedata *Expressionist* for Mass Spectrometry software, PCA models were generated with clear differences between the sample groups using non-discriminant analysis (Figure 3). From the combined data set, correlation networks were built which grouped the metabolites with similar profiles together, not only by analyte group, but by sample group and also by analysis type (Figure 4).

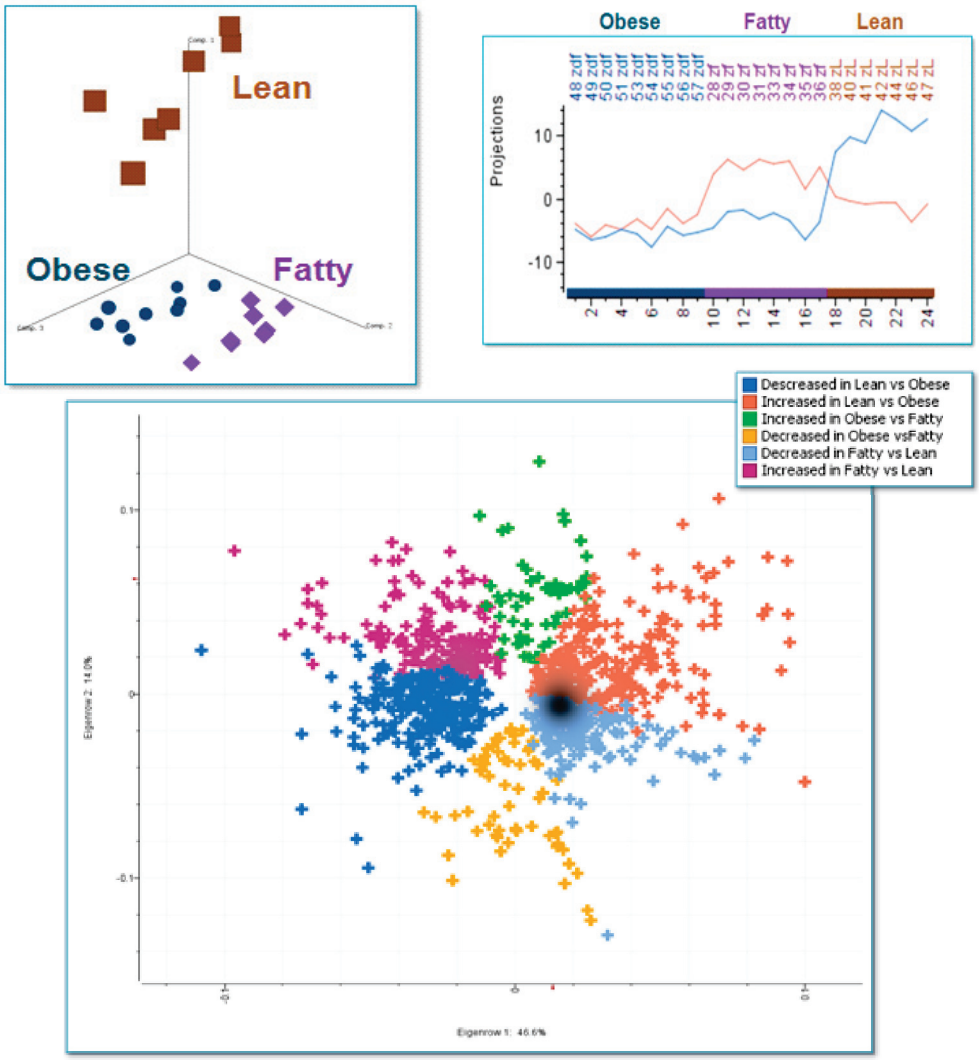


Figure 3. Principal Component Analysis (PCA) after ANOVA and Clustering. Top left figure shows an unsupervised PCA scores plot with sample groups represented as different shape and color. The diagram highlights that there are differences between the groups of samples. The lean (control) group clusters further away than the other two groups which provides confidence that the data is of good quality. Top right highlights the profiles of the top two eigenvectors responsible for the largest variation in the dataset. The bottom figure is the loadings plot from the top two eigenvectors which highlight the ions changing and in which groups. This plot is colored by the inclusion in the up/down regulation as determined by the contrast analysis with the metabolites found in the ANOVA testing.

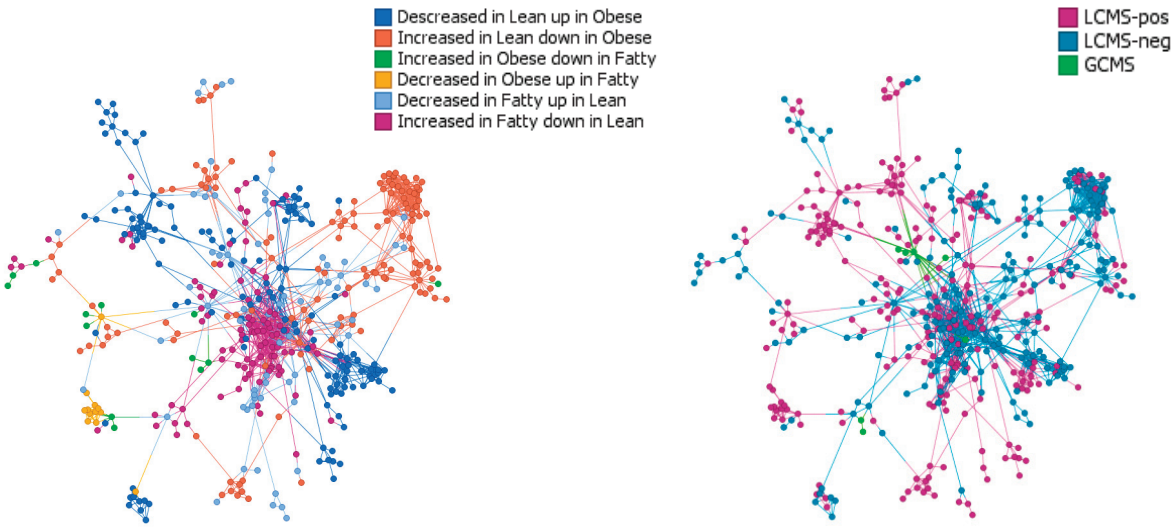


Figure 4. Visualizing the Correlation Networks. (Left) This view highlights the metabolites and their up/down regulation in each phenotype versus another phenotype. (Right) This figure shows all metabolites amenable to positive or negative mode LC/MS or GC/MS. It is apparent that there are areas of the network where one technique does not reach, but a combination of these two technologies allows complete coverage. The tight cluster in the middle makes visualization of many complementary findings between GC and LC analyses. When selected, nodes become highlighted with red outline and labeled with metabolite name.

Correlation networks are defined by a set of nodes which in this case are the metabolites. These nodes are connected by edges which refer to any direct or indirect correlation or interaction. Thus, a network can be visualized as a set of points connected by lines (Figure 4). Correlation coefficients that are calculated from the intensity values in the data are used for determining the pairwise attractive forces. Since the network is based on the correlation, the interpretation is performed in the context of metabolites that show correlation (in measured intensity values). Metabolites that are grouped together have similar correlation. The links between nodes in separate groups represent metabolites that have the next closest correlation.

Complementary Data Provides Additional Biological Insight

It is known that different metabolites are more amenable to either GC/MS or LC/MS, and that some metabolites can be observed by both techniques (Figure 1). This was confirmed in our data, where some metabolites such as the amino acids were detected in both the LC/MS and GC/MS analyses, but there were other classes of metabolites which were only detected using either of the chromatographic approaches. The overlap gave an extra level of confidence to our biological interpretation and validated our results. We also observed that while GC/MS addresses primary metabolites such as the TCA cycle and glycolysis metabolites, LC/MS most readily addressed secondary metabolites such as the glycerolipids and carnitines. This provided deeper, more informative coverage of the metabolome.

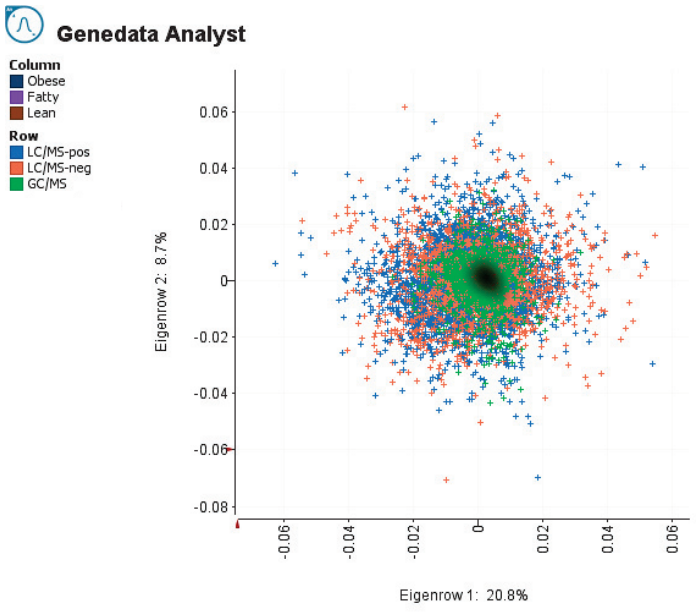


Figure 5. Loadings Plot Highlighting Variables (Ions) Detected by Each Analysis. The variables in green in the center of the plot are the primary ions detected by GC/MS, whereas the variables in blue and red on the outskirts are the secondary ions detected by LC/MS. If we only used one technology then we would miss data and impact our biological interpretation as pieces of our puzzle would be missing. Combining data from the LC/MS and GC/MS high resolution platforms allows a deeper understanding and broader coverage.

Conclusions

The overall objective of global metabolomics is to be able to detect as many metabolites as possible in the shortest possible time to enable a larger number of samples to be analyzed per study. Covering the chemical diversity of these metabolites is a challenging proposition and cannot be done using one single analytical technique such as GC or LC alone. Combining the strengths of each of these chromatographic approaches means that we are able to obtain a broader coverage of the metabolome and extract more valuable biological information. The full value of this approach can only be realized using a data processing strategy that enables the integration of both data types into a single analysis.

The combination of the *Pegasus* GC-HRT system (LECO) and the *TripleTOF* 5600 LC/MS system (AB SCIEX) with the *Expressionist* for Mass Spectrometer software (Genedata) provides a more comprehensive coverage of the metabolome and a deeper understanding of the biology.

- The three phenotypes were easily differentiated using unsupervised PCA on the combined dataset.
- Groups of analytes that showed significant up- or down-regulation between the phenotypes were identified.
- Correlation network analysis highlights how the data comes together to form a more comprehensive picture of the underlying biology of the Zucker rat model.

References

[1]Lei, Z., Huhman, D. V. and Sumner, L. 2011. Mass Spectrometry Strategies in Metabolomics. *The Journal of Biological Chemistry,* **286,** 25435-25442.

[2]Lin, Z., Zhang, Z., Lu, H., Jin, Y., Yi, L. and Liang, Y. 2014. Joint MS-based platforms for comprehensive comparison of rat plasma and serum metabolic profiling. *Biomedical Chromatography*, **28 (9)**, 1235-1245.

Pegasus and LECO are registered trademarks of LECO Corporation;
TripleTOF is a trademark of AB Sciex;
Expressionist is a registered trademark of Genedata.

Delivering the Right Results

Life Science and Chemical Analysis Solutions

LECO Corporation | 3000 Lakeview Avenue | St. Joseph, MI 49085 | Phone: 800-292-6141 | Fax: 269-982-8977
info@leco.com • www.leco.com | ISO-9001:2008 | HQ-Q-994 | *LECO is a registered trademark of LECO Corporation.*

© 2014 LECO Corporation